

# Accessibility percolation on n-trees

S. NOWAK<sup>1</sup> and J. KRUG<sup>1</sup>

<sup>1</sup> *Institute for Theoretical Physics, University of Cologne, D-50937 Köln, Germany*

PACS 64.60.ah – Percolation

PACS 02.50.Cw – Probability theory

PACS 87.23.Kg – Dynamics of evolution

**Abstract** – Accessibility percolation is a new type of percolation problem inspired by evolutionary biology. To each vertex of a graph a random number is assigned and a path through the graph is called accessible if all numbers along the path are in ascending order. For the case when the random variables are independent and identically distributed, we derive an asymptotically exact expression for the probability that there is at least one accessible path from the root to the leaves in an  $n$ -tree. This probability tends to 1 (0) if the branching number is increased with the height of the tree faster (slower) than linearly. When the random variables are biased such that the mean value increases linearly with the distance from the root of the tree, a percolation threshold emerges at a finite value of the bias.

**Introduction and outline.** – Percolation theory in its modern form was introduced by Broadbent and Hammersley in 1957 [1]. Since then percolation has become a cornerstone of probability theory and statistical physics, with applications ranging from molecular dynamics to star formation [2,3], and new variants of the problem continue to attract much attention (e.g., [4–7]). Standard percolation theory is concerned with the loss of global connectivity in a graph when vertices or bonds are randomly removed, as quantified by the probability for the existence of an infinite cluster of contiguous vertices.

Here we consider a novel kind of percolation problem inspired by evolutionary biology. Imagine a population of some lifeform endowed with the same genetic type (genotype). If a mutation occurs, a new genotype is created which can die out or replace the old one. Provided natural selection is sufficiently strong, the latter only happens if the new genotype has larger fitness. As a consequence, on longer timescales the genotype of the population takes a path through the space of genotypes along which the fitness is monotonically increasing [8]. Such a path is called selectively accessible [9,10].

Since the relationship between genotype and fitness is very complicated and largely unknown, it is natural to assign fitness values to genotypes in a random way. Evolutionary accessibility thus becomes a statistical property of random fitness landscapes [11–13]. In recent years it has become possible to experimentally determine fitness land-

scapes comprising small numbers of genetic loci [10,14,15], and the study of evolutionary accessibility is an important tool for characterizing and interpreting such data sets [12,16].

On an abstract level, the problem of interest is defined as follows: Consider a graph  $G$  where a continuous random variable  $w(\sigma)$  is assigned to each vertex  $\sigma$ . A path of contiguous vertices

$$\sigma_1 \rightarrow \sigma_2 \rightarrow \dots \rightarrow \sigma_n$$

through the graph is called accessible if

$$w(\sigma_1) < w(\sigma_2) < \dots < w(\sigma_n).$$

Accessibility percolation studies the statistics of such accessible paths, specifically the probability for the existence of paths that span the entire graph. An important difference to standard percolation is that accessibility depends on the discrete gradient of a globally defined fitness function  $w: G \mapsto \mathbb{R}$ , rather than on local random variables [13]. To see this, it is useful to formulate the problem as a kind of bond percolation on directed graphs, where the edge from  $\sigma$  to  $\sigma'$  is removed if  $w(\sigma) > w(\sigma')$  and a path is called accessible if all edges are present along the path. In the simple case when the  $w(\sigma)$  are chosen as independent and identically distributed (i.i.d.) random variables, this means that a given edge is removed with probability  $1/2$ , but in contrast to standard bond percolation the re-

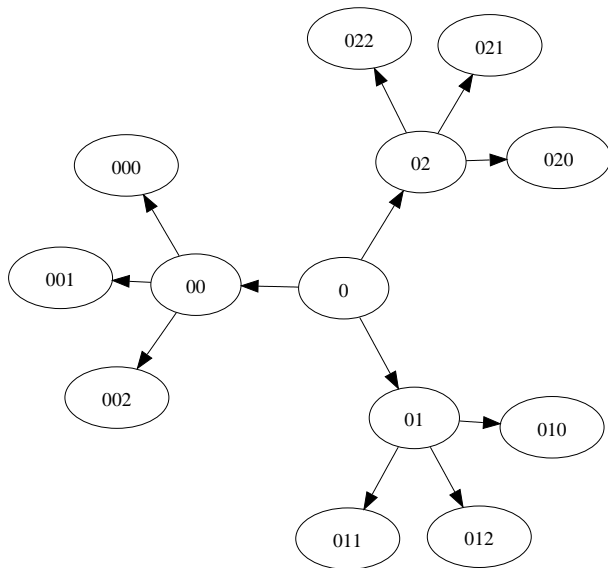


Fig. 1: An  $n$ -tree with branching number  $n = 3$  and height  $h = 2$ .

removal of different edges is correlated through the fitness function.

In the biological context the natural choice for the underlying graph  $G$  is the  $L$ -dimensional hypercube, where  $L$  is the number of binary genetic loci [11, 17–19]. The topology of the hypercube is rather peculiar in that the diameter of the graph (as measured by the length of the longest path) is equal to its dimension (as measured by the coordination number of a vertex). Recent numerical and analytic work has shown that accessibility percolation on the hypercube with i.i.d. random variables is *critical*, in the sense that minor changes of the model can trigger a transition from low to high accessibility [12, 20].

In order to further explore the role of graph topology in the phenomenology of accessibility percolation, we here consider the problem on  $n$ -trees. The simple structure of the trees allows us to obtain precise analytic bounds on the probability of existence of accessible paths. Moreover, by scaling the branching number of the tree with its height we are able to mimic the structure of the hypercube and place previous results [12, 20] into a broader context.

An  $n$ -tree or complete  $n$ -ary tree is a rooted tree where each vertex is connected to  $n$  child vertices, except for the leaves which have no children. The height  $h$  is defined as the distance from the root to the leaves, i.e., any path from the root to a leaf consists of  $h + 1$  vertices (cf. fig. 1). This graph is very similar to graphs known as Cayley trees or Bethe lattices, though these are usually defined such that each vertex has the same coordination number rather than the same number of child vertices, i.e., the root would have  $n + 1$  child vertices.

Our results are as follows. First we assume that the random variables  $w(\sigma)$  are drawn independently from a single continuous distribution, a setting known in the evolutionary context as the House of Cards (HoC) model [17, 21].

As accessibility is determined solely by the ordering of the random variables on the tree, in this case the results are independent of the choice of the underlying probability distribution. We will give an exact expression for the second moment of the number  $N$  of accessible paths from the root to the leaves. Furthermore, we will show that the probability  $\mathbb{P}(N \geq 1)$  that there is at least one accessible path is asymptotically equivalent to  $n^h/h!$  for large  $h$ . This result is valid if  $n$  is constant as well as if  $n = n(h)$  is a function that grows more slowly than linear with  $h$ . If the growth is faster than linear, the probability  $\mathbb{P}(N \geq 1)$  tends to 1 for  $h \rightarrow \infty$ . For a linearly growing function  $n(h) = \alpha h$  there is a percolation threshold, i.e.,  $\mathbb{P}(N \geq 1) \rightarrow 0$  for  $\alpha < \alpha_c$  and  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) > 0$  for  $\alpha > \alpha_c$  with  $\alpha_c \in [e^{-1}, 1]$ .

We then consider the effect of a bias that increases the mean fitness linearly with the distance from the root, corresponding to the Rough Mount Fuji (RMF) model of fitness landscapes in evolutionary biology [12, 16, 22, 23]. Specifically, we set  $w(\sigma) = x_\sigma - \theta d(\sigma)$ , where  $d(\sigma)$  is the distance to the closest leaf,  $\theta > 0$  is the bias parameter, and the  $x_\sigma$  are i.i.d. random variables which (for technical reasons to be explained below) are drawn from a Gumbel probability density function. We show that this model displays a percolation threshold at a nonzero value  $\theta_c$  of the bias, such that  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) = 0$  for  $\theta < \theta_c$  and  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) > 0$  for  $\theta > \theta_c$ .

**Basic properties of  $n$ -trees.** – Let us recall some basic properties of  $n$ -trees and their accessibility. The number of leaves is given by  $n^h$ . Since each leaf corresponds uniquely to a path, this is also the number of paths. In previous work on accessibility percolation on the hypercube, it has usually been assumed that the value of the destination vertex is the global fitness maximum [12, 20]. Analogously we will assume here that the starting vertex, i.e., the root, is the global minimum. For the case of i.i.d. random variables  $w(\sigma)$ , there are then  $h!$  different equally likely orderings of the values encountered along a path, and hence the probability that a given path is accessible is  $1/h!$ . By linearity the first moment of the number  $N$  of accessible paths follows immediately [12],

$$\langle N \rangle = \frac{n^h}{h!}. \quad (1)$$

Another useful property of  $n$ -trees is their recursive structure. A tree of height  $h + 1$  can be thought of as a root which is connected to  $n$  trees of height  $h$ . This property enables us to give an explicit but recursive equation for the probability of having no paths from the root to a leaf. Let us denote the probability that there is no path in a tree of height  $h + 1$ , given that the root has the value  $w(0) = x$ , by  $Q_h(x)$ . A tree of height  $h + 1$  is not accessible if each vertex  $\sigma$  with distance 1 to the root has either a value  $w(\sigma) < x$  or it has a larger value but the subtree of height

$h$  with  $\sigma$  as the root is not accessible. This leads to

$$Q_1(x) = F_0(x)^n$$

$$Q_{h+1}(x) = \left[ F_h(x) + \int_x^\infty f_h(y) Q_h(y) dy \right]^n \quad (2)$$

where  $f_d(x)$  and  $F_d(x)$  are the probability density function and cumulative distribution function, respectively, of the values  $w(\sigma)$  belonging to the vertices which have distance  $d$  to the closest leaf. For future reference we allow these distributions to depend explicitly on  $d$ . The probability of having at least one accessible path in a tree of height  $h$  is then given by

$$\mathbb{P}(N \geq 1) = 1 - \int_{-\infty}^\infty f_h(x) Q_h(x) dx. \quad (3)$$

So far we did not find a closed analytic solution of the recurrence relation (2). However, the equation can be solved numerically by approximating the integral by the trapezoidal rule. In the following we will use this to support and supplement our analytic results, which are based on the analysis of the moments of  $N$ .

**Calculation of the second moment.** – To compute the second moment of  $N$  we define indicator variables  $\Theta_i$  for each path  $i \in \{1, 2, \dots, n^h\}$  by

$$\Theta_i = \begin{cases} 1, & \text{if the } i\text{-th path is accessible,} \\ 0, & \text{else.} \end{cases} \quad (4)$$

The  $\Theta_i$  are dependent random variables but identically distributed with

$$\langle \Theta_i \rangle = \langle \Theta_i^2 \rangle = \mathbb{P}(\Theta_i = 1) = \frac{1}{h!}.$$

This yields  $N = \sum_{i=1}^{n^h} \Theta_i$  and hence

$$\langle N^2 \rangle = \sum_{i=1}^{n^h} \langle \Theta_i^2 \rangle + \sum_{\substack{i,j=1 \\ i \neq j}}^{n^h} \langle \Theta_i \Theta_j \rangle = \langle N \rangle + \sum_{\substack{i,j=1 \\ i \neq j}}^{n^h} \langle \Theta_i \Theta_j \rangle. \quad (5)$$

Note that the correlator  $\langle \Theta_i \Theta_j \rangle$  depends only on the number of vertices which the  $i$ -th and  $j$ -th path have in common. For any two paths  $i$  and  $j$  (with  $i \neq j$ ) which share  $h - k + 1$  vertices,  $\langle \Theta_i \Theta_j \rangle$  is given by the probability  $\pi_k$  that both paths are accessible. If the value  $w(\sigma)$  of the vertex  $\sigma$  where both paths diverge is given by  $x$ , the two paths can be decomposed into three independent subpaths (cf. fig. 2). All vertices which are closer to the root than  $\sigma$  must have a smaller value than  $x$  and all vertices closer to the leaves must have a larger value. Additionally, all values have to be in ascending order on each subpath. This yields

$$\pi_k = \int_0^1 \frac{x^{h-k-1}}{(h-k-1)!} \left( \frac{(1-x)^k}{k!} \right)^2 dx \quad (6) \quad \text{where}$$

$$= \frac{B(h-k, 2k+1)}{(h-k-1)! k!^2} = \binom{2k}{k} \frac{1}{(h+k)!}$$

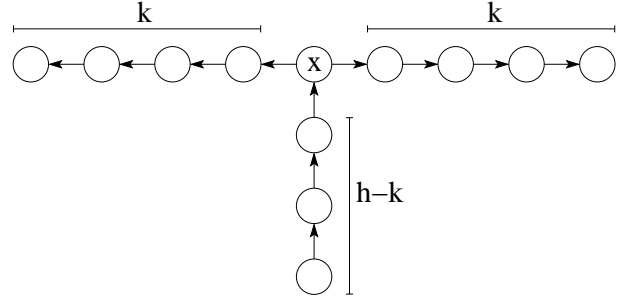


Fig. 2: The correlation between two paths only depends on the number  $h - k + 1$  of vertices both paths have in common.

where  $B(x, y)$  is Euler's beta function. To evaluate the sum over the correlators in eqn. (5) one also needs the number  $m_k$  of pairs of paths that have  $h - k + 1$  common vertices. This number can be evaluated with a simple combinatorial consideration: For the first path, say the left one in fig. 2, any leaf can be chosen, i.e., there are  $n^h$  possibilities. The second path shares  $h - k + 1$  vertices, then there are  $n - 1$  potential child vertices to choose from (if one takes the  $n$ -th vertex, it would also belong to the first path) and finally one can choose any child vertex until one reaches a second leaf which gives another  $n^{k-1}$  possibilities. Altogether there are  $m_k = (n - 1) n^{h+k-1}$  different pairs. This yields

$$\langle N^2 \rangle = \langle N \rangle + \sum_{k=1}^h \pi_k m_k$$

$$= \langle N \rangle + \frac{n-1}{n} \sum_{k=1}^h \binom{2k}{k} \frac{n^{h+k}}{(h+k)!} \quad (7)$$

$$\leq \langle N \rangle + \langle N \rangle^2 + \sum_{k=1}^{h-1} \binom{2k}{k} \frac{n^{h+k}}{(h+k)!}. \quad (8)$$

**Probability of having accessible paths.** – In order to gain information about the probability for the existence of accessible paths, we will use the first- and second moment method, i.e., we will apply the following lemma [20, 24]: Let  $X$  be a random variable which takes only integer values  $\{0, 1, 2, \dots\}$  and has a finite second moment. Then the inequality

$$\langle X \rangle \geq \mathbb{P}(X \geq 1) \geq \frac{\langle X \rangle^2}{\langle X^2 \rangle} \quad (9)$$

holds true. By combining this with eqns. (1) and (8) one obtains the inequality

$$\langle N \rangle \geq \mathbb{P}(N \geq 1) \geq \frac{\langle N \rangle^2}{\langle N \rangle + \langle N \rangle^2 + S(h)} \quad (10)$$

$$S(h) = \sum_{k=1}^{h-1} \binom{2k}{k} \frac{n^{h+k}}{(h+k)!}. \quad (11)$$

We will treat the branching number  $n$  as a function of  $h$ . Since we are going to distinguish the three cases of linear growth, growth faster than linear and growth slower than linear, we make the ansatz

$$n = n(h) = h \alpha(h). \quad (12)$$

The three cases correspond to  $\alpha(h) = \alpha = \text{const.}$ ,  $\alpha(h) \rightarrow \infty$  and  $\alpha(h) \rightarrow 0$ , respectively. Furthermore, we will assume that  $n(h+1) \geq n(h)$ . Using Stirling's formula we obtain for the mean value of  $N$

$$\langle N \rangle \approx \frac{n(h)^h}{\sqrt{2\pi h} h^h e^{-h}} = \frac{[e \alpha(h)]^h}{\sqrt{2\pi h}}. \quad (13)$$

This expression goes to zero if  $\alpha(h) \rightarrow 0$  or  $\alpha(h) \equiv \alpha \leq e^{-1}$  and diverges if  $\alpha(h) \rightarrow \infty$  or  $\alpha(h) = \alpha > e^{-1}$ .

First we consider the case  $\alpha(h) \rightarrow 0$  and show that  $S(h)$  grows more slowly than  $\langle N \rangle$ . Stirling's formula applied to eqn. (11) leads to

$$\begin{aligned} S(h) &\leq \sum_{k=1}^{h-1} \frac{4^k e^{k+h+1} n^{k+h}}{2\pi \sqrt{\pi k (k+h)} (k+h)^{k+h}} \\ &\leq \frac{n^h}{\sqrt{2\pi(h+1)(h+1)^h e^{-h}}} \sum_{k=1}^{h-1} \left( \frac{4en}{h+1} \right)^k \\ &\leq \langle N \rangle x \frac{1-x^{h-1}}{1-x} \quad \text{with} \quad x = \left( \frac{4en}{h+1} \right). \end{aligned}$$

Since  $x \rightarrow 0$  for large  $h$  also  $S(h) \langle N \rangle^{-1}$  tends to zero. This implies according to eqn. (10) that  $\mathbb{P}(N \geq 1)$  is asymptotically equivalent to  $\langle N \rangle$ , and both quantities vanish asymptotically for large  $h$  (cf. fig. 3).

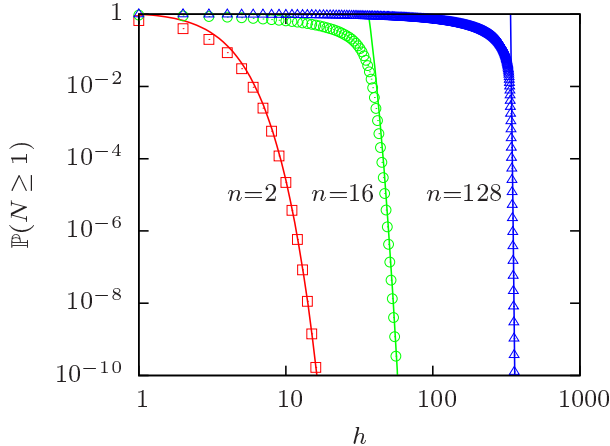


Fig. 3: The probability  $\mathbb{P}(N \geq 1)$  as a function of  $h$  for constant branching numbers  $n$ . Solid lines represent the asymptotic expression  $\mathbb{P}(N \geq 1) \sim \langle N \rangle$ . Symbols correspond to the numerical solution of the recurrence relation (2).

The case  $\alpha(h) \rightarrow \infty$  is a bit more complicated. Similar to the previous case, we will show that  $S(h)$  grows more slowly than  $\langle N \rangle^2$ , which implies that  $\mathbb{P}(N \geq 1) \rightarrow 1$

for  $h \rightarrow \infty$ . Instead of using Stirling's formula we will estimate the function  $\xi(h) = S(h) \langle N \rangle_h^{-2}$  recursively:

$$\begin{aligned} \xi(h+1) &= \frac{(h+1)!^2}{n(h+1)^{h+1}} \sum_{k=1}^h \binom{2k}{k} \frac{n(h+1)^k}{(h+k+1)!} \\ &\leq \frac{(h+1)^2}{(2h+1)n(h+1)} + \frac{(h+1)^2}{(h+2)n(h+1)} \xi(h) \\ &\leq \frac{1+\xi(h)}{\alpha(h)}. \end{aligned} \quad (14)$$

Since  $\alpha(h) \rightarrow \infty$ , we can fix any constant  $\Omega > 1$  and find  $h_0$  such that  $\alpha(h) > \Omega$  for all  $h > h_0$ . Then it can be shown by induction that

$$\xi(h) \leq \frac{\xi(h_0)}{\Omega^{h-h_0}} + \sum_{k=1}^{h-h_0} \frac{1}{\Omega^k} \xrightarrow{h \rightarrow \infty} \frac{1}{\Omega-1}$$

and hence  $\xi(h) \rightarrow 0$  since  $\Omega$  can become arbitrarily large.

Finally we consider the case  $\alpha(h) = \alpha = \text{const.}$  Using the recurrence relation (14), it can be shown by induction that  $\xi(h) \leq (\alpha-1)^{-1}$ , given that  $\alpha > 1$ . Hence the limit of  $\xi(h)$  is finite and according to eqn. (10) and (13) we get

$$\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) \begin{cases} = 0, & \text{for } \alpha \leq e^{-1}, \\ \geq 1 - \alpha^{-1}, & \text{for } \alpha > 1. \end{cases}$$

It is reasonable to assume that the limit is monotonic in  $\alpha$ . Therefore, there must be a threshold  $\alpha_c$  in  $[e^{-1}, 1]$  such that  $\mathbb{P}(N \geq 1) \rightarrow 0$  for  $\alpha < \alpha_c$  and  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) > 0$  for  $\alpha > \alpha_c$ .

**Effect of a bias.** — In this section we assume that the fitness values  $w(\sigma)$ , rather than being i.i.d., are distributed according to  $f(x + d(\sigma)\theta)$  where  $d(\sigma)$  is the distance from  $\sigma$  to the closest leaf and  $f(x) = \exp(-x - \exp(-x))$  is the probability density of the Gumbel distribution. The reason for this choice is that for this distribution the probability for the random variables along a path to be in ascending order (the ‘ordering probability’) is known in closed form [23]. Using this result it follows that the probability for a path of length  $k$  directed towards the leaves to be accessible is

$$P_k(\theta) = a(\theta)^k b(\theta, k), \quad (15)$$

where

$$a(\theta) = 1 - e^{-\theta} \quad \text{and} \quad b(\theta, k) = \left[ \prod_{m=1}^k (1 - e^{-m\theta}) \right]^{-1}. \quad (16)$$

Note that  $b(\theta, k)$  is monotonically increasing with  $k$  and converges for  $k \rightarrow \infty$ . This leads to the bounds

$$1 \leq b(\theta, k) \leq \lim_{k \rightarrow \infty} b(\theta, k) =: b(\theta) \quad (17)$$

with [23]

$$b(\theta) = \sqrt{\frac{\theta}{2\pi}} \exp\left(\frac{\pi}{6\theta} - \frac{\theta}{24}\right). \quad (18)$$

In the following we will show that  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) = 0$  for  $a(\theta) < n^{-1}$  and  $\lim_{h \rightarrow \infty} \mathbb{P}(N \geq 1) > 0$  for  $a(\theta) > n^{-1}$ , and therefore the percolation threshold occurs at  $\theta_c = \ln(n) - \ln(n-1)$ .

First let  $\theta < \theta_c$ , i.e.,  $a(\theta) < n^{-1}$ . Then according to eqn. (9) we get

$$\mathbb{P}(N \geq 1) \leq \langle N \rangle = P_{h+1}(\theta) n^h \leq a(\theta) b(\theta) [a(\theta) n]^h \rightarrow 0$$

for  $h \rightarrow \infty$ .

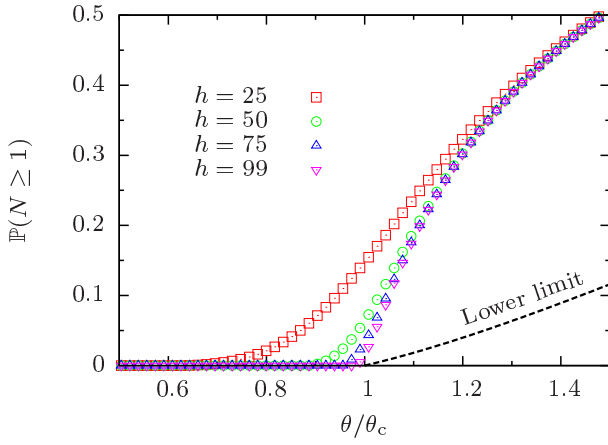


Fig. 4: Probability of finding accessible paths in a 4-tree as a function of the bias  $\theta$ . The critical bias is given by  $\theta_c = \ln(4/3) \approx 0.288$ . Symbols represent the numerical solution of eqn. (2), the lower bound is given by eqn. (19).

Now let  $\theta > \theta_c$ . In order to apply eqn. (9) one needs to compute the second moment which in turn requires the calculation of the probability  $\pi_k$  that two paths that share  $h-k+1$  vertices are both accessible. A necessary condition for this is that the  $h-k+1$  common vertices are accessible as well as both separated subpaths which consist of  $k$  vertices each (cf. fig. 2). Therefore  $\pi_k \leq P_{h-k+1}(\theta) P_k(\theta)^2$  and hence

$$\begin{aligned} \sum_{\substack{i,j=1 \\ i \neq j}}^{n^h} \langle \Theta_i \Theta_j \rangle &= \sum_{k=1}^h m_k \pi_k \leq \sum_{k=1}^h m_k P_{h-k+1}(\theta) P_k(\theta)^2 \\ &= (n-1) \sum_{k=1}^h n^{h+k-1} a(\theta)^{h+k+1} b(\theta, h-k+1) b(\theta, k)^2. \end{aligned}$$

After using the upper bound for  $b(\theta, k)$  from eqn. (17), introducing  $y = na(\theta) > 1$  and evaluating the geometric sum one gets

$$\sum_{\substack{i,j=1 \\ i \neq j}}^{n^h} \langle \Theta_i \Theta_j \rangle \leq \frac{b(\theta)^3}{n} \frac{y^{h+2}(y^h - 1)}{y - 1}$$

which finally leads to

$$\begin{aligned} \mathbb{P}(N \geq 1) &\geq \frac{\langle N \rangle^2}{\langle N^2 \rangle} = \frac{\langle N \rangle}{1 + \langle N \rangle^{-1} \sum_{i \neq j} \langle \Theta_i \Theta_j \rangle} \\ &\geq \frac{a(\theta) b(\theta, h+1) y^h}{1 + b(\theta)^2 (y^h - 1) y (y - 1)^{-1}} \\ &\xrightarrow{h \rightarrow \infty} \frac{a(\theta)}{b(\theta)} \frac{y - 1}{y} > 0. \end{aligned} \quad (19)$$

The lower bound proves that there is a threshold, but the numerical solution of eqn. (2) indicates that it is not a good estimate for the exact value of  $\mathbb{P}(N \geq 1)$  (cf. fig. 4). Nevertheless the transition is clearly continuous, in qualitative agreement with the behavior of the bound (19).

Note that the threshold criterion  $a(\theta_c) = n^{-1}$  applies whenever the probability that a path of length  $k$  is accessible can be written as in eqn. (15), where  $b(\theta, k)$  is bounded by two positive functions of  $\theta$ , i.e.  $c_1(\theta) \geq b(\theta, k) \geq c_2(\theta) > 0$  for all  $k$  and  $\theta > 0$ . In particular, for  $b(\theta, k) = 1$  and  $a(\theta) = \theta$  it reduces to the case of standard percolation where vertices are removed with probability  $1 - \theta$  and a path is accessible if all vertices are present. This leads directly to the well-known result  $\theta_c = n^{-1}$  for standard percolation in  $n$ -trees [2]. A bound on  $P_k$  of the form (15) has been obtained in [20] for RMF models with a fairly general class of probability distributions, indicating that the behavior described here is not restricted to the Gumbel distribution.

**Conclusions.** – In this paper we studied evolutionary accessibility as a new type of percolation problem and considered the statistics of accessible paths on trees. For the case of i.i.d. random variables, we presented a method for the exact calculation of the second moment  $\langle N^2 \rangle$  of the number of accessible paths. Based on this result, we could show by the second moment method that the probability  $\mathbb{P}(N \geq 1)$  of finding accessible paths is asymptotically equivalent to the mean value  $\langle N \rangle$  for large heights  $h$  and constant branching number  $n$ . When  $n$  is scaled with  $h$ , we show that  $\mathbb{P}(N \geq 1)$  goes to 0 or 1 depending on whether  $n(h)$  grows more slowly or faster than linearly. In case of a linear scaling  $n \propto h$  the behavior depends on the constant of proportionality  $\alpha$ . We can show that the limit of  $\mathbb{P}(N \geq 1)$  is positive if  $\alpha > 1$  and zero if  $\alpha \leq e^{-1}$ , but our method fails for values in between.

The tree with linear scaling  $n = \alpha h$  is of particular interest, as it approximates the topology of the hypercube. There are different ways of relating the two graphs that lead to different values of the prefactor  $\alpha$ . Observing that the coordination number of the hypercube is equal to its diameter one could set  $n = h$  and hence  $\alpha = 1$ . On the other hand, the total number of pathways traversing a hypercube of size  $h$  is  $h! \sim (h/e)^h$ , which can be equated to the number of paths on the tree,  $n^h$ , to yield  $\alpha = e^{-1}$ . In either case, the comparison to the tree shows that the hypercube is poised near the threshold between high and low accessibility. This is consistent with the rigorous



analysis of [20], which shows that the HoC model on the hypercube undergoes an abrupt transition from  $\mathbb{P}(N \geq 1) \approx 1$  to  $\mathbb{P}(N \geq 1) \approx 0$  as the fitness value of the *starting* vertex is increased.

The different topological structure of hypercubes and trees also manifests itself in the response to a bias  $\theta$  applied to the random variables along the paths. On the hypercube any finite bias  $\theta > 0$  leads to high accessibility,  $\mathbb{P}(N \geq 1) \approx 1$  [12, 20], while on the tree with fixed branching number  $n$  a continuous percolation transition occurs at a threshold value  $\theta_c > 0$ . Although the explicit calculation of the threshold was restricted here to the case of a Gumbel distribution, we expect this behavior to apply for a large class of distributions (see [20, 23]).

Our results as well as the inequality (9) indicate that the probability of having at least one accessible path is, in general, closely related to their mean number. The left part of the inequality implies that the probability goes to zero if the mean does. If the mean diverges, a non-vanishing probability for accessible paths results provided the ratio  $\langle N^2 \rangle / \langle N \rangle^2$  remains bounded as the graph size tends to infinity. This was the case in the examples considered here, but it need not always be true. Explicit counterexamples are the HoC model on the hypercube without conditions on the starting vertex, for which  $\langle N \rangle = 1$  and  $\mathbb{P}(N \geq 1) \rightarrow 0$  [12, 20], and the block model of protein evolution [25], for which  $\langle N \rangle \rightarrow \infty$  and  $\mathbb{P}(N \geq 1) \rightarrow 0$  as the hypercube dimension  $L \rightarrow \infty$  [26]. It is clear from (9) that this behavior requires very large fluctuations in  $N$ , and it will be interesting to further explore these fluctuations and their impact on accessibility.

Obviously, accessibility percolation can also be studied on other graphs. But on most structures which are not tree-like it is quite complicated to calculate the number of paths between two points which is necessary to calculate the second moment and to apply the inequality (9). Nevertheless the problem could be analyzed numerically or with other analytical methods.

\* \* \*

This work was supported by DFG within SPP 1590 and the Bonn Cologne Graduate School of Physics and Astronomy. We thank Peter Hegarty and Anders Martinsson for useful correspondence, and Anton Bovier for discussions.

## REFERENCES

- [1] BROADBENT S. R. and HAMMERSLEY J. M., *Proc. Cambridge Philos. Soc.*, **53** (1957) 629
- [2] STAUFFER D. and AHARONY A., *Introduction to Percolation Theory* (CRC Press) 1994
- [3] SAHIMI M., *Applications of percolation theory* (CRC Press) 1994
- [4] WILKINSON D. and WILLEMSSEN J.F., *J. Phys. A: Math. Gen.*, **16** (1983) 3365
- [5] ADLER J., *Physica A*, **171** (1991) 453
- [6] ACHLIOPTAS D., D'SOUZA R.M. and SPENCER J., *Science*, **323** (2009) 1454
- [7] KNECHT C.L., TRUMP W., BEN-AVRAHAM D. and ZIFF R.M., *Phys. Rev. Lett.*, **108** (2012) 045703
- [8] GILLESPIE J.H., *Evolution*, **38** (1984) 1116
- [9] WEINREICH, D. M., WATSON, R. A. and CHAO, L., *Evolution*, **59** (2005) 1165
- [10] WEINREICH D. M., DELANEY N. F., DEPRISTO M. A. and HARTL D. L., *Science*, **312** (2006) 111
- [11] CARNEIRO M. and D. L. HARTL, *PNAS*, **107** (2010) 1747
- [12] FRANKE J., KLÖZER A., DE VISSER J. A. G. M. and KRUG J., *PLoS Comput. Biol.*, **7** (2011) e1002134
- [13] FRANKE J. and KRUG J., *J. Stat. Phys.*, **148** (2012) 705
- [14] LUNZER M., MILLER S. P., FELSHEIM R. and DEAN A. M., *Science*, **310** (2005) 499
- [15] POELWIJK F. J., KIVIET D. J., WEINREICH D. M. and TANS S. J., *Nature*, **445** (2007) 383
- [16] SZENDRO I.G., SCHENK M.F., FRANKE J., KRUG J. and DE VISSER J.A.G.M., *J. Stat. Mech.*, (2013) P01005
- [17] KAUFFMAN S. and LEVIN S., *J. Theor. Biol.*, **128** (1987) 11
- [18] GAVRILETS S. and GRAVNER J., *J. Theor. Biol.*, **184** (1997) 51
- [19] GOKHALE C. S., IWASA Y., NOWAK M. A. and TRAULSEN A., *J Theor Biol.*, **259** (2009) 613
- [20] HEGARTY P. and MARTINSSON A., arXiv:1210.4798 [math.PR] preprint (2012)
- [21] KINGMAN J. F. C., *J. Appl. Prob.*, **15** (1978) 1
- [22] AITA T., UCHIYAMA, H., INAKA, T., NAKAJIMA, M., KOKUBO, T. and HUSIMI, Y., *Biopolymers*, **54** (2000) 64
- [23] FRANKE J., WERGEN G. and KRUG J., *J. Stat. Mech.*, (2010) P10013
- [24] ALON N. and SPENCER J., *The Probabilistic Method*, 2nd Ed. (Wiley) 2000
- [25] PERELSON A.S. and MACKEN C.A., *Proc. Natl. Acad. Sci. USA*, **92** (1995) 9657
- [26] SCHMIEGELT B., *Bachelor thesis* (University of Cologne) 2012